

Examen de statistique exploratoire du 16 décembre 2019

Tous documents écrits autorisés, téléphones et ordinateurs portables interdits, les calculatrices ne sont pas utiles,

Le sujet contient des annexes fournies dans un document séparé

Exercice 1

Vous pouvez répondre à chaque question en 2 ou 3 phrases et/ou 2 ou 3 formules.

1. Donner deux échantillons de taille 5 ayant la même médiane mais pas la même moyenne.
2. On dispose de 2 variables observées simultanément sur les mêmes 100 individus, on note a et b les vecteurs de taille 100 contenant les données correspondantes. On observe que la covariance entre a et b est de 1. Que peut on en conclure?
3. Par quelle règle détermine-t-on si un point est atypique quand on construit un boxplot?
4. En ACP comment définit on la contribution globale d'un individu à l'inertie totale du nuage des individus?

Exercice 2

On s'intéresse aux performances des 28 athlètes présents à l'épreuve de décathlon des JO d'Athènes de 2004. Les 10 épreuves sont la course sur 100m, saut en longueur, lancer de poids, saut en hauteur, course sur 400m, course 110m haies, lancer de disque, saut à la perche, lancer du javelot, course sur 1500m. Les noms de ces variables dans les tables et sorties sont respectivement *c100m Longueur Poids Hauteur c400m c110mhaies Disque Perche Javelot c1500m*. De plus on a une variable *Nom* contenant les noms des athlètes. On note les résultats des athlètes à chaque épreuve, mesurés en secondes pour les courses, et en mètres pour les sauts et les lancers, et on les écrit dans un tableau individus \times variables (donc avec 28 lignes et 10 colonnes) qui est donné en Table 1.

Partie ACP

On réalise une **ACP normée** sous SAS de ce tableau. On a considéré que tous les individus ont le même poids, à savoir $1/28$.

1. D'après les données de la table 1, qui est l'athlète qui a eu la meilleure performance en course sur 100 mètres? Et qui est l'athlète qui a eu la meilleure performance en saut en longueur? D'après la table 2, comment interpréter vous le fait que les variables "course sur 100 mètres" et "saut en longueur" soient nettement corrélées négativement?
2. Comment justifiez vous le fait d'avoir choisi une ACP normée (et non pas canonique)?

3. Quelle est l'inertie totale du nuage des individus (justifiez la réponse)?
4. D'après la table 3, combien d'axes principaux doit on utiliser pour interpréter les résultats de l'ACP si l'on utilise la règle de Kaiser (justifiez la réponse)?
5. D'après la table 4, quels sont les 2 individus qui contribuent le plus à la formation de l'axe principal 1 et qui sont projetés du côté positif de cet axe? Même question pour le côté négatif.
6. D'après la table 5, quelle est la variable initiale la plus corrélée avec la 1ère composante principale, et quelle est la variable la plus anti-corrélée (c'est à dire ayant la corrélation la plus négative) avec la 1ère composante principale? Justifiez la réponse.
7. Y a-t-il des variables mal représentées dans le premier **plan** factoriel?
8. Donner une interprétation simultanée de la carte des individus (figure 1) et de la carte des variables (figure 2) fournies par l'ACP, uniquement pour la première dimension.

Partie Classification

On réalise une **Classification Ascendante Hiérarchique** sous SAS. On a considéré encore que tous les individus ont le même poids, à savoir $1/28$, et à nouveau on a travaillé sur les données centrées et réduites.

1. On s'intéresse à l'étape de l'algorithme à l'issue de laquelle on obtient 16 classes. D'après la table 6, quels sont les individus ou classes qui ont été réunis lors de cette étape? Quelles sont l'inertie totale, interclasse et intraclasse à l'issue de cette étape?
2. D'après la figure 4 et la figure 5, quel nombre de classes proposeriez vous de retenir? Justifiez votre réponse.
3. Par la suite on a choisi arbitrairement la classification en 3 classes (attention ce n'est pas forcément le choix guidé par les critères de la figure 4 et la figure 5). D'après la table 7 et la table 8, donner une description de la composition des classes obtenues. Voyez vous des liens avec les résultats de l'ACP précédente?
4. A votre avis, qui est monté sur le podium olympique de l'épreuve du décathlon lors des jeux olympiques d'Athènes de 2004?